researchvision.us

Symmetria : Journal of Advanced Physics and Mathematical Review
Vol.3 | No. 1 | 2026 pp. 16–25 |ISSN: XXXX-XXXX

Statistical Modeling of Rainfall in Iraq: A Comparative Study of Classical and Flexible Probability Distributions

Iman Abdel Ali
Department of Statistics, University of Kirkuk, Kirkuk, Iraq
Email: dr.emanali@uokirkuk.edu.iq

Abdulqader Ahmed Jasim
Department of Statistics, University of Kirkuk, Kirkuk, Iraq
Email: abdulqaderahmed@uokirkuk.edu.iq

Wakaa Ali Hadba
Department of Statistics, University of Kirkuk, Kirkuk, Iraq
Email: wakaa2017@uokirkuk.edu.iq

Submitted: 18-Dec, 2025
Accepted: 08-Jan, 2026
Published: 14-Feb, 2026

Vol. 3, No. 1, 2026.
10.62762/SJAPMR.2024.000000

***Corresponding author:**
Iman Abdel Ali [1]

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

http://creativecommons.org/licenses/by/4.0/

Open Access

Abstract

Probabilistic modeling of rainfall data is an effective framework for analyzing the stochastic behavior of hydrological processes. This research involved fitting several distributions to the observed rainfall data in Iraq including the Gamma, Lognormal, Weibull, Generalized Weibull, GEV, and Nakagami distributions. The maximum likelihood method was used for parameter estimation. The adequacy of the model was evaluated through a series of goodness-of-fit statistics which include coefficient of determination, root mean square error, information criteria as well as the Kolmogorov–Smirnov test. The analysis shows the Gamma distribution performed best overall, with Weibull and GEV following closely behind, while Lognormal and Generalized Weibull were clearly poor fits to the observed rainfall data. Understanding streamflow distributions is vital for predicting streamflow behavior and improving hydrological forecasts. Furthermore, proper quantification can be useful for ecological management and mitigation efforts. When considering statistics, the research proves the impact of the Gamma distribution in modeling positively skewed rainfall. Thus, this can be used to conduct further stochastic analysis in applied hydrology.

Keywords: Rainfall modeling, Probability distributions, Maximum likelihood estimation, Goodness-of-fit tests, Model selection, Hydrological statistics

Introduction

Rainfall is an important hydrological variable affecting the availability of water resources, agricultural production and social resilience, especially in arid and semi-arid regions like Iraq.

The irregular pattern of daily rainfall cause frequent droughts as well as flash floods. Appropriate modeling is needed for managing water resources as well as for climate-adaptive strategies (Kottegoda & Rosso, 2008)

Statistical modeling of rainfall has a long-established past, based on classical probability theory. Thom (1958) was one of the first to apply the Gamma distribution to monthly precipitation (any atmospheric water including rainfall). This computationally set the stage for indices such as the Standardized Precipitation Index (SPI). Gumbel (1958) continued the pioneering work on formal extreme value theory and Generalized Extreme Value (GEV) distribution. Building on this theoretical foundation, Coles (2001) offers a general framework for modeling extremal behavior, while Kottegoda & Rosso (2008) goes on to extend it to stochastic time series methods in hydrology. Together, these contributions underlie rainfall statistics modelling nowadays.

Since 2010, we have seen lots of research in places with different weather. According to El Adlouni et al. (2015) and Modarres (2010), GEV and Gamma distributions were utilized to model the most extreme precipitation in arid climates. In Pakistan, Hussain et al. (2017) compared two- and three-parameter models and found that in semi-arid conditions, the Weibull and Lognormal distributions usually outperform the Gamma model. In Turkey, Dabanlı et al. (2020) showed that Generalized Weibull was the best model to yield skewed results. Further, Kaya & Dabanlı (2024) validated Nakagami and GEV as good models in the Mediterranean regimes. Kumar et al. (2024) also agreed with these findings in South Asia and suggested flexible distribution to assess rainfall risk.

When it comes to the Iraqi context, most studies explicitly focus on rainfall and not precipitation in general. According to Al-Ansari et al. (2012) and Al-Ansari (2014, 2019) estimated the Gamma and Lognormal models in Iraq north. Rasheed and Al-Mukhtar (2015) applied Weibull and Gumbel fits to the extreme rainfall data in Baghdad and found a high skewness. The works of Ahmed et al. (2019) and Al-Timimi & Jassim (2020) introduced probabilistic modeling to the central and southern Iraq. applying aridity indicators, Mohammed (2020) examined rainfall variability in a few Iraqi stations. Al-Obaidi (2020) investigated the seasonal and yearly rainfall characteristics in Daqouq District using GIS techniques.In Mosul rainfall analysis, Mohammed & Aziz (2021) evaluated the quality of Gamma and Lognormal fits. Meanwhile, Hussein et al. (2022) validated monthly Rainfall fits using more distributions, including Nakagami and GEV. According to Abbas and Ali (2023) who worked in western Iraq, flexible probability distributions capture the intermittent nature of rainfall better. Multi-decadal rainfall was studied in Kirkuk by Hussein & Kasim (2024). They noted that performance of GEV and Nakagami were better than Gamma and Lognormal for extremes. Also, long-term rainfall of the Kurdistan Region was seen in Fawzee et al. (2020). They remarked that Gamma and Weibull fit best for seasonal variation and performance of Lognormal is poor for skewed conditions.

While this work has progressed, there has not been a systematic prolonged examination of different probability distributions, including old and modern flexible ones, for monthly rainfall in Iraq. To bridge this gap, we assess six candidate distributions: Gamma, Lognormal, Weibull, Generalized Weibull, GEV and Nakagami. The monthly Rainfall data (1981-2024) is used for

estimating parameters through maximum likelihood estimation (MLE) and assessing the performance of the models using a range of goodness-of-fit measures (AIC, BIC, Kolmogorov-Kolmogorov–Smirnov, Anderson-Darling, RMSE and R2). Thus, this study seeks to discover the model that fits the rainfall data in Iraq statistically most appropriate. A suitable model is important for forecasting, drought modelling, and water resource planning.

Materials and Methods

This section describes the rainfall dataset and the statistical methods used for modelling. Several distributions Gamma, Lognormal, Weibull, Generalized Weibull, GEV, and Nakagami were selected to model rainfall characteristics.

The parameters were estimated using maximum likelihood estimation (MLE), with model adequacy assessed through several goodness-of-fit tests.

Results and Discussion

The rainfall dataset utilized in this study ranges from 1981 to 2024, covering a 44-year period, and was obtained from the Iraqi Meteorological Organization and Seismology (IMOS). This is the annual total rainfall (mm) recorded from the different climatic conditions of the research period. The dataset is also subject to extremely dry and wet years which is good for statistical modelling. Table 1 presents descriptive statistics on rainfall data, which reveals a minimum of 0.78 mm, a maximum of 82.31 mm, a mean of 22.50 mm, and a standard deviation of 19.42 mm. The distributional features show that the distribution has moderate positive skewness of 0.53 and kurtosis of 2.35 implying a non-normal distribution of rainfall. The dataset serves as the basis for probabilistic modeling in this study, allowing multiple statistical distributions to be evaluated in order to better define rainfall occurrence.

Table1: Descriptive statistics of Rainfall data for used locations

Descriptive statistics	Value (mm)
Minimum	0.78
Maximum	82.31
Mean	22.5
Standard deviation	19.42
Skewness	0.53
Kurtosis	2.35

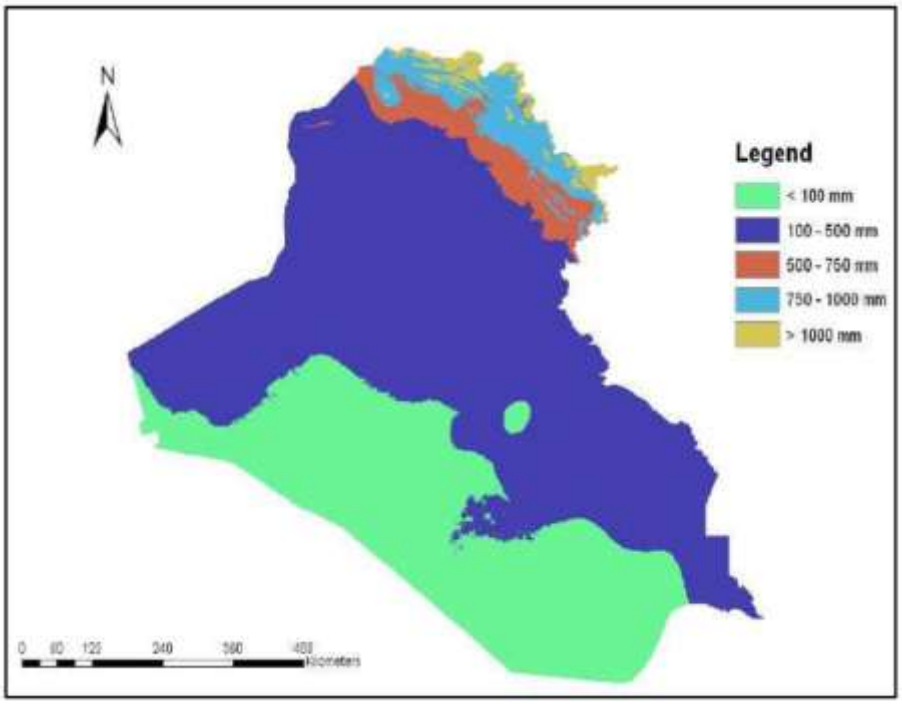


Figure 1. Annual Rainfall Map of Iraq, highlighting north-south gradient.

Method

Gamma Distribution

The Gamma distribution is a widely used continuous distribution of a random variable that takes on only positive values and has a right skew. A common use of the Gamma distribution is to describe rainfall amounts. Its probability density function is given by.:

$$f(x;k,\theta) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}} \; ; x>0, \; \theta, k>0 \qquad (1)$$

In the equation stated, k is the shape parameter which is greater than 0 and θ is the scale parameter which is greater than 0. Furthermore, Bs e-θx is the probability density function. Additionally, (k) is the Gamma function. The flexibility of this distribution makes it appropriate for the variability and asymmetry of rainfalls..

The Lognormal Distribution

Lognormal distribution is useful for positive-valued and skewed variables. Similarly, rainfall data is often skewed and high, which makes it a good candidate for this distribution. The log of the rainfall amounts is normally distributed as per this distribution. Its probability density function (PDF) is given by.

$$f(x;\mu,\sigma) = \frac{1}{x\,\sigma\sqrt{2\pi}} \; e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \; ; x>0, \sigma>0 \qquad (2)$$

The symbols '?' and 'σ' indicate the meanlocated at the standarddeviation scaleofvariable xi. The Lognormal distribution has proven successful in estimating moderate to extreme rainfall and is widely used in hydrology and climatology studies because it can effectively model skewed data..

Weibull Distribution

The Weibull distribution was first introduced by Waloddi Weibull (1951) and it is a flexible

tool widely employed in hydrology, climatology, and reliability. This model can capture increasing, decreasing, and constant hazard rates. Therefore it is used to model rainfall and wind speed. Weibull distribution probability density function (PDF) is given by.

$$f(x;k,\theta) = \frac{k}{\theta^k} x^{k-1} e^{-(\frac{x}{\theta})^k} \; ; x > 0, \; \theta, k > 0 \qquad (3)$$

where k > 0 is the shape parameter, θ > 0 is the scale parameter.

Generalized Weibull (GW) Distribution

The GW distribution was conceptualized by Mudholkar, Srivastava, and Kollia (1996) which is a generalized version of classical Weibull distribution with more flexibility to model lifetime and environmental data. Many experts in reliability, survival analysis and hydrology have recently become interested in the new distributions listed below. The probability density function (PDF) is given by:

$$f(x;\alpha,\beta,\gamma) = \frac{\gamma\beta}{\alpha} (\frac{x}{\alpha})^{\alpha-1} \left(1 - e^{-(\frac{x}{\alpha})^\beta}\right)^{\gamma-1} e^{-(\frac{x}{\alpha})^\beta}; x > 0, \; \alpha, \beta, \gamma > 0 \qquad (4)$$

Here $\alpha > 0$ is the scale parameter and $\beta > 0$, $\gamma > 0$ are shape parameters. When $\gamma = 1$, the model transforms into the usual Weibull distribution, and it takes on various shapes of the hazard rate, above all the "bathtub" shape. Because it is flexible, it can also be used to model rainfall and other hydrological processes.

Generalized Extreme Value (GEV)

The GEV distribution was created through the Extreme Value Theory that focuses on the asymptotic behavior of block maxima (and minima) of random variables. The GEV introduced and formalized by Jenkinson (1955) consists of three types of extreme value distributions: Gumbel (type I), Fréchet (type II), and Weibull (type III), unified by one shape parameter. Because of this, it is a basic tool for hydrology, climatology and risk analysis. In particular extreme rainfall modelling, floods and wind speed.

GEV distribution has probability density function (PDF).

$$f(x;\mu,\sigma,\xi) = \frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1} \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \qquad (5)$$

$$1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0, \quad \sigma > 0$$

for x satisfying $1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0$,with location $\mu \in R$, scale $\sigma > 0$, and shape $\xi \in R$.

Nakagami Distribution

The Nakagami distribution is proposed by Nakagami (1960) for fading modeling in radio wave propagation. Since that time, it has found extensive applications in hydrology, climatology, and environmental sciences and due to its flexibility to model data which is positively skewed, it has been applied in environmental and rainfall studies.

The Nakagami distribution uses a probability density function.

$$f(x;\mu,\Omega) = \frac{2}{\Gamma(\mu)} (\frac{\mu}{\Omega})^\mu \chi^{2\mu-1} e^{-\frac{\mu}{\Omega}\chi^2}; x > 0 \qquad (6)$$

where $m > \frac{1}{2}$ is the shape parameter, $\Omega > 0$ is the scale parameter, and.

Parameter Estimation

In the study, all candidate probability distribution parameters were estimated by MLE

(Maximum Likelihood Estimation) method. MLE, or maximum likelihood estimation, is known to be one of the most reliable techniques for parametric estimation (Casella & Berger, 2002). Because of its consistency, efficiency and asymptotic normality, it is widely used in statistics. This method involves constructing the likelihood function of the rainfall data as per some distribution. Thus, it represents the joint probability of the observed data. This function can be maximized to get parameter estimates.

$L(\theta)= \prod_{i=1}^{n} f(x_i;\theta)$ (7)

and the maximum likelihood estimator is obtained as:

$\hat{\theta}=\arg\max_{\theta} L(\theta)$ (8)

Because rainfall is skewed and heavy-tailed in oceanographic and hydrological applications, MLE is a good estimator. This method produces parameter estimates that are stable and represent the central location as well as extremes. The estimates we ended up with were used in the goodness fit test to evaluate the correctly fit probability that fit our rainfall data..

Goodness of Fit Test Criteria

To check the goodness-of-fit of fitted probability distributions, we applied several well-known criteria that are mostly used, including coefficient of determination (R^2), root mean square error (RMSE), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and finally Kolmogorov–Smirnov (KS) test with its p-value.

The coefficient of determination is defined as.:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{F}_i-\bar{F})^2}{\sum_{i=1}^{n}(\hat{F}_i-\bar{F})^2+\sum_{i=1}^{n}(F_i-\hat{F}_i)^2} \quad (9)$$

where F_i and $\hat{F}_i$ denote the empirical and theoretical cumulative probabilities, respectively, and $\bar{F}$ is their mean**.**

The RMSE is computed as:

$$\text{RMSE} = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(F_i-\hat{F}_i)^2\right)} \quad (10)$$

The information criteria are expressed as:

$\text{AIC} = -2\ \ln L\ (\hat{\theta}) + 2k$ (11)

$\text{BIC} = -2\ \ln L\ (\hat{\theta}) + k \ln n$ (12)

where $L\ (\hat{\theta})$ is the maximum likelihood function,k is the number of estimated parameters, and n is the sample size.

The KS test statistic is given by:

$$D_n = \sup_x \left|F_n(x) - \hat{F}(x;\theta)\right| \quad (13)$$

where $F_n(x)$ is the empirical cumulative distribution function and $\hat{F}(x;\theta)$ is the theoretical CDF with estimated parameters. The corresponding $\text{KS}_{\text{p-value}}$ is also reported to assess the statistical significance of the test results.

Results and Discussion

The parameter estimation results for the considered probability distributions, obtained using the Maximum Likelihood Estimation (MLE) method, are presented in Table 2. The Gamma distribution was characterized by a shape parameter of 2.0824 and a scale of 2.7082 (equivalent

to a rate of 0.3692). The Lognormal distribution exhibited a shape (sdlog) of 0.7911 and a mean log of 1.4709. For the Weibull distribution, the estimated parameters were a shape of 1.5221 and a scale of 6.2746. The Generalized Weibull distribution showed a shape of 0.0385, a scale of 3.8795, and an additional parameter c = 1. In the case of the GEV distribution, the parameters included shape = 0.1335, scale = 2.6145, and location = 3.7535. Finally, the Nakagami distribution yielded estimates of m = 0.6716 and ω = 46.8373.

Table 2. Estimated parameters of the probability distributions.

Distribution	Shape	Scale	Location
Gamma	2.082	2.708	–
Lognormal	1.574	0.842	–
Weibull	1.522	6.275	–
Generalized Weibull	0.998	4.462	1.220
GEV	-0.115	3.746	5.821
Nakagami	0.672	46.837	–

The goodness-of-fit statistics for all candidate distributions are presented in Table 3, based on multiple evaluation criteria including the coefficient of determination (R^2), root mean square error (RMSE), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Kolmogorov–Smirnov (KS) test with corresponding p-values.

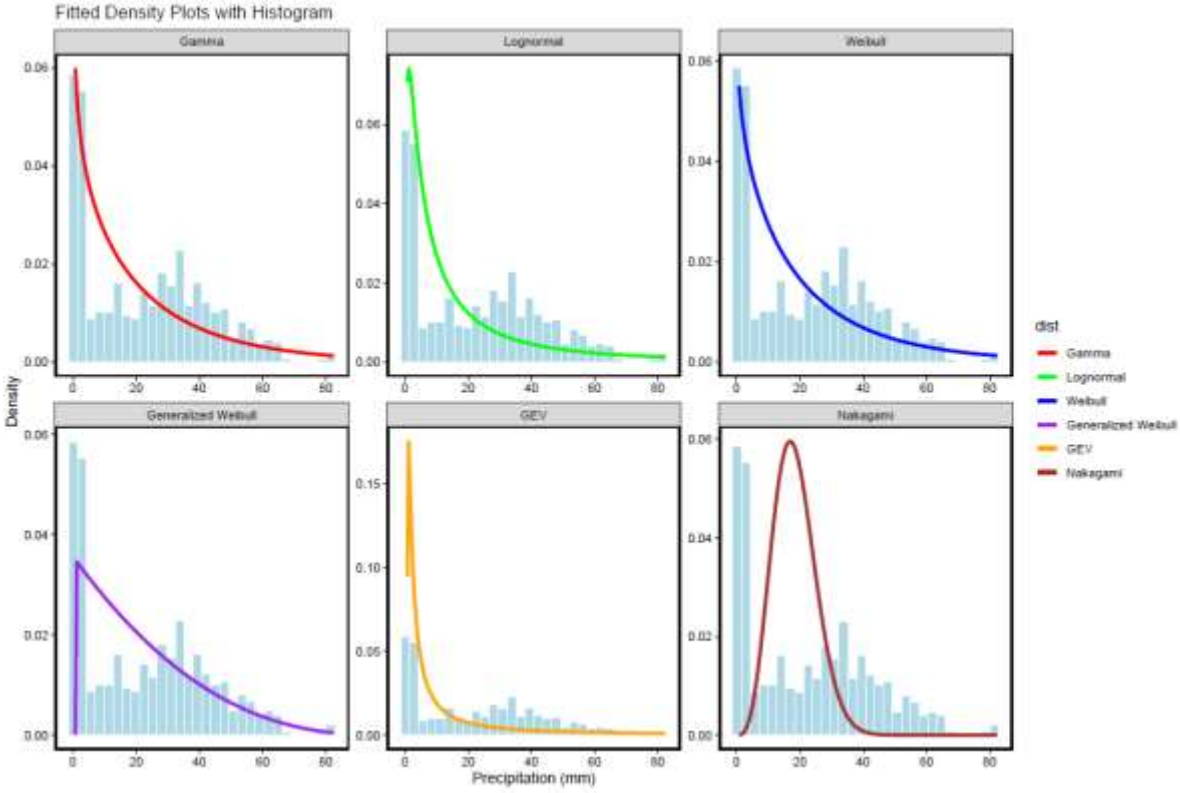
Table 3. Goodness-of-fit statistics for the fitted probability distributions.

Distribution	R^2	RMSE	AIC	BIC	KS	KS p-value
Gamma	0.999	0.136	5208.140	5217.960	0.026	0.497
Lognormal	0.932	2.028	5315.060	5324.880	0.072	0.000
Weibull	0.994	0.319	5218.770	5228.580	0.036	0.149
Generalized Weibull	0.982	5.626	7605.100	7619.830	0.670	0.000
GEV	0.983	0.567	5240.480	5255.200	0.027	0.480
Nakagami	0.984	0.528	5246.400	5256.220	0.059	0.002

Of all models considered, the Gamma distribution showed the best performance overall, with the highest R^2 value (0.9988) and the lowest RMSE value (0.1355). Also, it had the smallest AIC (5208.144) and BIC (5217.959) values. Moreover, the KS statistic for the Gamma model was 0.0262; the KS p-value is 0.496. We cannot reject the null hypothesis of the KS test (data follows the Gamma distribution) at 5% level. Thus, this is the best model. The second best model is Weibull (R2 = 0.9937, RMSE=0.3194, KS p-value=0.149. Hence, this is also a good model though slightly worse than Gamma. The GEV distribution performed well as well, especially in estimating extremes, R^2 = 0.983 and KS p-value = 0.480. On the other hand, Gamma distribution is the best probability model to describe the annual rainfalls of Iraq for the period (1981-2024).

Furthermore, its performance is better than Nakagami, GEV, Weibull, Lognormal and Generalized Weibull models. Also, Generalized Weibull model has the worst performance with a very high RMSE (5.6264), largest AIC and BIC values as well as a KS p-value of 0 indicating a very poor fit. The finding is in accordance with previous works on hydrology in arid and semi-arid areas as the gamma distribution tends to display more flexibility than wnge.

Figure 2 shows the probability you can assign to the amount of rain that falls in a given month. The Gamma and the Weibull distribution fits the observed data better as compared to the other distributions, especially centering and upper values. Lognormal distribution fit the extreme value fairly well, however, overestimation in the tail occurred. The Generalized Weibull provided a little more flexibility with a moderate fit. The GEV distribution is recommended to apply for climate extremes rather than overall distribution fitting, even though it fits extremes. On the other hands, the Nakagami distribution deviated from observed rainfall clearly especially at higher values confirming its goodness of fit was poor.



Conclusion

The study tested the fitness of six probability distributions for annual rainfall in Iraq covering a period of 1981 - 2024 which includes Gamma, Weibull, Lognormal, Generalized Weibull, GEV and Nakagami. The parameters were estimated using the Maximum Likelihood Estimation (MLE) method and the goodness-of-fit was tested by R^2, RMSE, AIC, BIC and the Kolmogorov–Smirnov (KS) test. The results show that the Gamma distribution outperforms the other candidate models with the highest R^2 value of 0.999, lowest RMSE (0.136) and the lowest AIC and BIC values. The KS test supports that the Gamma distribution is the best fit for the rainfall

distribution of Iraq; the p-value is 0.497. The Weibull distribution was the second best fit while the Generalized Weibull and Nakagami distribution performed poorly indicating unsuitability for modeling Rainfall in this case. The results show that the Gamma distribution is the best model for rainfall variability in Iraq, corroborating findings in studies of arid and semi-arid areas. The Gamma distribution can be efficiently used for hydrological modeling, water resources planning, and climate impact assessment in Iraq. For future research, stationary distribution models could be applied in conjunction with climate covariates to detect potential trends and variance in a modified climate.

References

1. Abbas, H., & Ali, K. (2023). Modeling precipitation extremes in Iraq using flexible probability distributions. Theoretical and Applied Climatology, 152(3–4), 1123–1137.
2. Ahmed, A., Al-Kubaisi, Q., & Al-Obaidi, A. (2019). Analysis of precipitation variability in Iraq. Journal of Water and Climate Change, 10(3), 613–625.
3. Al-Ansari, N. (2012). Managing water resources in Iraq: The long road ahead. International Journal of Water Resources Development, 28(4), 795–817.
4. Al-Ansari, N. (2014). Water resources in Iraq: Perspectives and prognoses. Engineering, 6(8), 667–684.
5. Al-Ansari, N. (2019). Hydro-politics of the Tigris and Euphrates. In M. M. Aral (Ed.), Water resources in the Middle East (pp. 199–220). Springer.
6. Al-Obaidi, A. S. (2020). GIS-based geographical analysis of seasonal and annual rainfall characteristics in Daqouq District. Kirkuk University Journal for Humanitarian Studies, 17(1), 163–181.
7. Al-Timimi, Y. M., & Jassim, H. (2020). Statistical analysis of rainfall trends in Iraq. Environmental Earth Sciences, 79(17), 446.
8. Casella, G., & Berger, R. L. (2002). Statistical inference (2nd ed.). Duxbury Press.
9. Coles, S. (2001). An introduction to statistical modeling of extreme values. Springer.
10. Dabanlı, İ., Şen, Z., & Yilmaz, A. O. (2020). Modeling precipitation series in Turkey with various probability distributions. Theoretical and Applied Climatology, 139(3–4), 1221–1234.
11. El Adlouni, S., Ouarda, T. B. M. J., et al. (2015). Generalized maximum likelihood estimators for the generalized extreme value distribution. Journal of Hydrology, 521, 478–490.
12. Fawzee, M. A., Salh, S. M., & Ahmed, S. A. (2020). Statistical distribution of rainfall in Kurdistan-Iraq region. Al-Mustansiriyah Journal of Science, 30(4), 8–28.
13. Gumbel, E. J. (1958). Statistics of extremes. Columbia University Press.
14. Hussain, M., Pasha, G. R., & Ahmad, M. (2017). Goodness-of-fit tests for rainfall data in Pakistan. Pakistan Journal of Statistics and Operation Research, 13(1), 37–46.
15. Hussein, A. H., & Kasim, M. N. (2024). Utilizing statistical distribution tests to develop rainfall intensity–duration–frequency curves for enhanced hydrological analysis in Kirkuk city, Iraq. Water Practice & Technology, 19(11), 4378–4389.
16. Hussein, Z., Ali, S., & Hassan, R. (2022). Rainfall variability and probability distribution

fitting in northern Iraq. SN Applied Sciences, 4(2), 115.
17. Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly Journal of the Royal Meteorological Society, 81(348), 158–171.
18. Kaya, Y., & Dabanlı, İ. (2024). Assessment of probability distributions for daily precipitation modeling in Mediterranean climates. Atmospheric Research, 294, 107–125.
19. Kottegoda, N. T., & Rosso, R. (2008). Applied statistics for civil and environmental engineers (2nd ed.). Wiley-Blackwell.
20. Kumar, R., Singh, V. P., & Yadav, B. (2024). Stochastic modeling of rainfall using flexible probability distributions. Hydrological Sciences Journal, 69(2), 245–259.
21. Modarres, R. (2010). Regional frequency distribution type of annual precipitation in arid regions of Iran. Water Resources Management, 24(10), 2471–2485.
22. Mohammed, F., & Aziz, A. (2021). Evaluation of probability distributions for annual rainfall in Iraq. Arabian Journal of Geosciences, 14, 1215.
23. Mohammed, J. H. (2020). Calculation of aridity index and determination of climate type for rainy months in selected stations in Iraq. Kirkuk University Journal for Scientific Studies, 15(2), 63-78.
24. Mudholkar, G. S., Srivastava, D. K., & Kollia, G. D. (1996). A generalization of the Weibull distribution with application to the analysis of survival data. Journal of the American Statistical Association, 91(436), 1575–1583.
25. Nakagami, M. (1960). The m-distribution—A general formula of intensity distribution of rapid fading. In W. C. Hoffman (Ed.), Statistical methods in radio wave propagation (pp. 3–36). Pergamon Press.
26. Rasheed, S., & Al-Mukhtar, S. (2015). Climate change impact on rainfall patterns in Iraq. Arabian Journal of Geosciences, 8(11), 9133–9144.
27. Thom, H. C. S. (1958). A note on the gamma distribution. Monthly Weather Review, 86(4), 117–122.
28. Weibull, W. (1951). A statistical distribution function of wide applicability. Journal of Applied Mechanics, 18(3), 293–297.